

Методология анализа причин ошибочной классификации фотоизображений программами компьютерного зрения на основе искусственного интеллекта в медицинской визуализации

А.И. Ламоткин, Д.И. Корабельников

Автономная некоммерческая организация дополнительного профессионального образования «Московский медико-социальный институт им. Ф.П. Гааза» (2-я Брестская ул., д. 5, Москва 123056, Российская Федерация)

Для контактов: Андрей Игоревич Ламоткин, e-mail: lamotkin.an@yandex.ru

РЕЗЮМЕ

Актуальность. Внедрение программ компьютерного зрения (ПКЗ) на основе технологий искусственного интеллекта (ИИ) в практику медицинской диагностики предъявляет принципиально новые требования к качеству входных фотоизображений. Условия фотосъемки в различных областях медицины существенно различаются, в связи с чем референсные диапазоны фотометрических и текстурных параметров, при которых модели ИИ обеспечивают воспроизводимую диагностическую точность, должны устанавливаться отдельно для каждой предметной области. Систематический анализ причин ошибочных заключений ПКЗ выступает необходимым условием их клинического применения и совершенствования.

Цель: разработка обобщенной методологии анализа причин ошибочной классификации фотоизображений ПКЗ на основе моделей ИИ, обеспечивающей установление референсных диапазонов фотометрических и текстурных параметров для конкретной области медицинской визуализации, а также формирование критериев исключения

Мы предоставляем данную авторскую версию для обеспечения раннего доступа к статье. Эта рукопись была принята к публикации и прошла процесс рецензирования, но не прошла процесс редактирования, верстки, присвоения порядковой нумерации и корректуры, что может привести к различиям между данной версией и окончательной отредактированной версией статьи.

We are providing this an author-produced version to give early visibility of the article. This manuscript has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the final typeset and edited version of the article.

изображений с аномальными значениями параметров при работе модулей предобработки в составе медицинских приложений.

Материал и методы. Методология включает восемь последовательных этапов. Сформирован набор фотоизображений с верифицированными заключениями на основании гистологических исследований и дерматоскопии, осуществлена их классификация ПКЗ и категоризация заключений по четырем стандартным категориям (истинно положительные, истинно отрицательные, ложноположительные, ложноотрицательные). Для каждого фотоизображения вычисляли набор из 13 фотометрических и текстурных метрик качества: яркость, контраст, резкость, энтропия, высокочастотная насыщенность, доли пересвеченных и затемненных пикселей, средние значения и стандартные отклонения цветовых каналов. Систематические межгрупповые различия выявляли посредством однофакторного дисперсионного анализа (англ. ANalysis Of VAriance, ANOVA), критерия Уэлча и корреляционного анализа по Спирмену. Локализацию областей фотоизображений, определяющих решение нейронной сети, осуществляли методами объяснимого ИИ. Референсные диапазоны устанавливали по характеристикам правильно классифицированных фотоизображений (категории истинно положительных и истинно отрицательных результатов) как интервалы $[\text{mean} - 2 \cdot \text{std}; \text{mean} + 2 \cdot \text{std}]$. Эффективность нормализации параметров оценивали по приросту точности, чувствительности и специфичности.

Результаты. На примере дерматологических ПКЗ Derma Onko Check и Melanoma Check показано, что предложенная методология позволяет выявить статистически значимые межгрупповые различия фотометрических и текстурных параметров ($F=13,50-39,31$ и $p<0,001$ для основных метрик ANOVA; $F=5,41-72,29$ и $p<0,001$ для категории заключения по ANOVA), подтвердить независимость закономерностей от программы ПКЗ ($p=0,39-0,96$ для фактора программы; $p=0,15-0,92$ для эффекта взаимодействия), верифицировать различие групп заключений ПКЗ по совокупности метрик (лямбда Уилкса 0,639; $F=10,37$; $p<0,001$) при многомерном анализе и установить ключевые независимые предикторы ошибочного заключения по логистической регрессии (размытие по быстрому преобразованию Фурье: отношение шансов (англ. odds ratio, OR) 3,08; резкость: OR 0,31; доля пересвеченных пикселей: OR 1,64). Установлены референсные диапазоны яркости (0,467–0,942), контраста (0,066–0,333), энтропии (3,626–5,590) и высокочастотной насыщенности (23,82–56,48), а также критические пороги (доля пересвеченных или затемненных пикселей более 55%), при превышении которых фотоизображение исключается из инференса. Применение модуля адресной предобработки, нормализующего отклоняющиеся параметры до референсных диапазонов, обеспечило

прирост диагностической точности на +0,014–0,017 в абсолютных значениях во всех исследованных конфигурациях ПКЗ при преимущественном повышении специфичности (+0,015–0,019).

Заключение. Предложенная методология анализа причин ошибочной классификации фотоизображений с помощью ПКЗ апробирована в ПКЗ Derma Onko Check и Melanoma Check на ограниченной выборке (460 фотоизображений, две морфологические подгруппы меланоцитарных опухолей кожи). Распространение методологии и установленных референсных диапазонов на другие области медицинской визуализации (офтальмологию, гистологию, ультразвуковую диагностику), где условия получения фотоизображений принципиально различаются, требует валидации на представительных мультицентровых выборках с пересчетом референсных диапазонов параметров в соответствии с условиями получения фотоизображений каждой предметной области. Реализация в составе программ модулей исключения изображений с аномальными значениями метрик является закономерным практическим следствием применения методологии и обеспечивает воспроизводимое повышение клинической точности систем ПКЗ на основе ИИ.

КЛЮЧЕВЫЕ СЛОВА

программы компьютерного зрения, искусственный интеллект, ошибки классификации, метрики качества, объяснимый искусственный интеллект, медицинская визуализация

Для цитирования

Ламоткин А.И., Корабельников Д.И. Методология анализа причин ошибочной классификации фотоизображений программами компьютерного зрения на основе искусственного интеллекта в медицинской визуализации. *ФАРМАКОЭКОНОМИКА. Современная фармакоэкономика и фармакоэпидемиология*. 2026; 19 [принятая рукопись]. <https://doi.org/10.17749/2070-4909/farmakoekonomika.2026.382>.

Methodology for analyzing the causes of photo image misclassification by computer vision programs based on artificial intelligence in medical imaging

A.I. Lamotkin, D.I. Korabelnikov

Moscow Haass Medical and Social Institute (5 2nd Brestskaya Str., Moscow 123056, Russian Federation)

Corresponding author: Andrey I. Lamotkin, e-mail: lamotkin.an@yandex.ru

ABSTRACT

Background. The introduction of artificial intelligence (AI)-driven computer vision programs (CVPs) into medical diagnostics has imposed stricter requirements on the quality of input photographic images. Imaging conditions vary significantly across medical fields, necessitating the establishment of field-specific reference ranges for photometric and textural parameters to ensure that AI models maintain reproducible diagnostic accuracy. A systematic analysis of the causes of erroneous classifications by CVPs is essential for their clinical application and further improvement.

Objective: To develop a generalized methodology for analyzing the causes of classification errors in photographic images by processed by AI-driven CVPs. The proposed methodology enables the establishment of reference ranges for photometric and textural parameters for specific medical imaging applications, as well as the development of criteria for excluding images with anomalous values when during preprocessing in medical software systems.

Material and methods. The methodology includes eight sequential stages. A dataset of photographic images verified by histological and dermatoscopic examinations was compiled, classified using the CVPs, and assigned to either of the four standard categories (true positive, true negative, false positive, and false negative). For each photographic image, thirteen photometric and textural quality metrics were calculated, including: brightness, contrast, sharpness, entropy, high-frequency saturation, proportions of overexposed and underexposed pixels, and mean values and standard deviations of the color channels. Systematic between-group differences were identified using one-way ANalysis Of VAriance (ANOVA), Welch's test, and Spearman's rank correlation analysis. The image regions that determine the neural network decision were localized using explainable AI techniques. Reference ranges were established from the characteristics of correctly classified photographs (true positive and true negative categories), defined as intervals of [mean - 2 std; mean + 2 std]. The effectiveness of parameter normalization was assessed by the improvement in accuracy, sensitivity, and specificity.

Results. The proposed methodology was tested using the Derma Onko Check and Melanoma Check CVPs as an example. Its application allowed statistically significant intergroup differences in photometric and textural parameters to be identified ($F=13.50-39.31$, $p<0.001$ for the main metrics of the one-way ANOVA; $F=5.41-72.29$, $p<0.001$ for the conclusion category by the two-way ANOVA). The analysis confirmed that the observed patterns were independent of the specific CVP used ($p=0.39-0.96$ for the program factor; $p=0.15-0.92$ for the interaction effect). Multivariate analysis further demonstrated significant differences among classification

outcome groups based on the combined set of image quality metrics (Wilks' lambda 0.639; $F=10.37$; $p<0.001$) and established key independent predictors of classification errors through logistic regression (Fast Fourier Transform blur: odds ratio (OR) 3.08; sharpness: OR 0.31; proportion of overexposed pixels: OR 1.64). Reference ranges were established for brightness (0.467–0.942), contrast (0.066–0.333), entropy (3.626–5.590), and high-frequency saturation (23.82–56.48), along with critical thresholds for image exclusion from inference (proportion of overexposed or darkened pixels greater than 55%). The use of a targeted preprocessing module for normalization of deviating parameters falling outside the reference ranges ensured an increase in diagnostic accuracy by +0.014–0.017 in absolute values across all studied CVP configurations, with a predominant increase in specificity (+0.015–0.019).

Conclusion. The proposed methodology for analyzing the causes of erroneous classification of photographic images by AI-driven CVPs was tested on the example of Derma Onko Check and Melanoma Check using a limited dataset (460 photographic images representing two morphological subgroups of melanocytic skin tumors). The extension of the methodology to other areas of medical imaging (ophthalmology, histology, or ultrasound diagnostics), where image acquisition conditions differ substantially, will require validation on representative multicenter datasets with recalculation of parameter reference ranges to reflect the imaging specifics of each domain. The integration of modules for detecting and excluding images with abnormal metric values constitutes a natural practical implication of the proposed methodology and ensures a reproducible increase in the clinical accuracy of AI-driven CVP systems.

KEYWORDS

computer vision program, artificial intelligence, classification errors, quality metrics, explainable artificial intelligence, medical imaging

For citation

Lamotkin A.I., Korabelnikov D.I. Methodology for analyzing the causes of photo image misclassification by computer vision programs based on artificial intelligence in medical imaging. *FARMAKOEKONOMIKA. Sovremennaya farmakoekonomika i farmakoepidemiologiya / FARMAKOEKONOMIKA. Modern Pharmacoeconomics and Pharmacoepidemiology*. 2026; 19 [accepted manuscript] (in Russ.). <https://doi.org/10.17749/2070-4909/farmakoekonomika.2026.382>.

| Основные моменты | Highlights |
|---|--|
| Что уже известно об этой теме? | What is already known about the subject? |
| Программы компьютерного зрения (ПКЗ) на основе искусственного интеллекта (ИИ) демонстрируют высокую диагностическую точность при анализе медицинских фотоизображений в дерматологии, офтальмологии, гистологии и других областях, сопоставимую с уровнем диагностической точности опытных врачей-специалистов | Artificial intelligence (AI)- driven computer vision programs (CVPs) demonstrate high diagnostic accuracy in analyzing medical photo images in dermatology, ophthalmology, histology, etc., comparable to the level of experienced medical specialists |
| Нейронные сети функционируют как «черные ящики», что затрудняет выявление причин ошибочных заключений (ложноположительных и ложноотрицательных) | Neural networks function as “black boxes”, making it difficult to identify the causes of erroneous conclusions (false positives and false negatives) |
| Качество входных фотоизображений (фотометрические и текстурные параметры) существенно влияет на точность работы моделей ИИ | The quality of input photographs (photometric and textural parameters) has a significant effect on the accuracy of AI models |
| Что нового дает статья? | What are the new findings? |
| Разработана обобщенная методология, состоящая из 8-этапного анализа причин ошибочной классификации, включающая вычисление 13 фотометрических и текстурных метрик, многоуровневый статистический анализ (ANOVA, MANOVA, логистическая регрессия) и методы объяснимого ИИ | A generalized, eight-step methodology for analyzing the causes of misclassification has been developed, including the calculation of 13 photometric and textural metrics, multilevel statistical analysis (ANOVA, MANOVA, logistic regression), and explainable AI methods |
| Установлены статистически значимые межгрупповые различия параметров между верными и ошибочными заключениями, ключевые независимые предикторы ошибок (размытие по быстрому преобразованию Фурье OR=3,08; резкость OR=0,31; доля пересвеченных пикселей OR=1,64) и референсные диапазоны параметров (например, яркость 0,467–0,942, энтропия 3,626–5,590 для меланоцитарных опухолей) | Statistically significant intergroup differences in parameters between correct and incorrect conclusions were identified, along with key independent predictors of errors (Fast Fourier Transform blur OR=3.08; sharpness OR=0.31; proportion of overexposed pixels OR=1.64), and reference ranges for parameters (e.g., brightness 0.467–0.942, entropy 3.626–5.590 for melanocytic tumors) |

| | |
|---|---|
| Разработан и валидирован модуль адресной предобработки, нормализующий отклоняющиеся параметры, который обеспечил прирост точности на +0,014–0,017 и специфичности на +0,015–0,019 при тестировании на программах ИИ Derma Onko Check и Melanoma Check | A preprocessing module was developed and validated for normalizing outlier parameters. This resulted in an increase in accuracy of +0.014–0.017 and specificity of +0.015–0.019 when tested using Derma Onko Check and Melanoma Check CVPs |
| Как это может повлиять на клиническую практику в обозримом будущем? | How might it impact the clinical practice in the foreseeable future? |
| Интеграция модулей адресной предобработки и автоматического исключения изображений с аномальными параметрами в медицинские приложения ИИ позволит существенно повысить воспроизводимость и клиническую надежность ПКЗ, особенно в телемедицине и первичном звене | Integrating modules for preprocessing and automatically excluding images with anomalous parameters into medical AI applications will significantly improve the reproducibility and clinical reliability of CVPs, particularly in telemedicine and primary care |
| Стандартизация условий фотосъемки и применение референсных диапазонов, специфичных для каждой предметной области (дерматология, офтальмология, гистология и др.), может снизить долю ошибочных заключений, обусловленных техническим качеством фотоизображений | Standardizing imaging conditions and using reference ranges specific to each subject area (dermatology, ophthalmology, histology, etc.) can reduce the rate of classification errors caused by the quality of photographic images |
| Методология станет основой для дальнейшего совершенствования моделей ИИ и разработки рекомендаций по качеству входных данных, что ускорит безопасное внедрение ИИ-инструментов в рутинную клиническую практику и улучшит качество диагностики новообразований кожи и других патологий | The proposed methodology forms a basis for further improvement of AI models and the development of recommendations for input data quality, accelerating the safe implementation of AI tools in routine clinical practice and improving the quality of diagnosis of skin neoplasms and other pathologies |

ВВЕДЕНИЕ / INTRODUCTION

Программы компьютерного зрения (ПКЗ) на основе технологий искусственного интеллекта (ИИ) в последние годы устойчиво занимают все более значимое место в медицинской диагностике. В различных областях клинической практики продемонстрирована принципиальная возможность применения таких систем для решения задач классификации и сегментации: в дерматологии – при дифференциальной диагностике новообразований кожи [1, 2], в офтальмологии – при анализе фотоизображений глазного дна [3], в гистологии – при обработке цифровых препаратов

[4], в эндоскопии и ультразвуковой диагностике – при выявлении патологических изменений [5]. Высокая диагностическая точность, сопоставимая в ряде исследований с точностью врачей-специалистов, открывает перспективы интеграции подобных инструментов в первичное звено здравоохранения и телемедицину.

Нейронные сети функционируют как «черные ящики» [6]: при возникновении ошибочных заключений (ложноположительных или ложноотрицательных) практически невозможно без специального анализа установить, чем именно они обусловлены – дефектами архитектуры модели, недостатками обучающей выборки или систематическими проблемами качества входных фотоизображений [7]. Это принципиально затрудняет как настройку алгоритмов разработчиками, так и формирование рекомендаций по стандартизации условий фотосъемки.

Принципиальное обстоятельство, требующее систематической проработки, состоит в том, что условия получения фотоизображений в различных областях медицинской визуализации различаются. Дерматологические фотоизображения формируются при макросъемке кожных покровов в условиях рассеянного видимого света с расстояния 8–15 см; фотоизображения глазного дна получают с помощью специализированных фундус-камер с инфракрасной подсветкой через расширенный зрачок; гистологические цифровые препараты обрабатываются на сканере микропрепаратов при многократном оптическом увеличении и стандартизованной окраске. Каждой из этих областей соответствуют собственные нормативные значения яркости, контраста, цветового распределения и текстурной информативности. Распространение референсных диапазонов одной области на другую методологически некорректно: перенос модели ИИ, обученной на дерматологических фотоизображениях, на фундус-фотоизображения без переустановки референсных диапазонов закономерно приведет к росту доли ошибочных заключений.

В ранее опубликованных работах проведен ретроспективный анализ ошибочных заключений программ *Derma Onko Check* и *Melanoma Check* при классификации немеланоцитарных опухолей кожи [8]. С применением методик в области объяснимого ИИ – *Score-CAM* (англ. *class activation mapping*)¹ и *Occlusion Sensitivity*², а также математической статистики установлены статистически значимые различия фотометрических и текстурных характеристик фотоизображений между группами истинных и ложных классификаций. Дополнительно проведенная статистическая оценка с помощью однофакторного и двухфакторного дисперсионного анализа (англ. *ANalysis Of*

¹ Метод картирования активации классов, в котором для взвешивания карт признаков используются не градиенты, а значения уверенности модели.

² Чувствительность к окклюзии, метод анализа, который помогает понять, какие части изображения или данных наиболее важны для классификации глубокой нейронной сети.

VAriance, ANOVA), многомерного дисперсионного анализа (англ. Multivariate ANalysis Of VAriance, MANOVA) и множественной логистической регрессии подтвердила и уточнила полученные закономерности на трех независимых уровнях, обеспечив одновременный учет категории заключения и применяемой ПКЗ, оценку совокупного вектора метрик и количественное ранжирование вклада отдельных параметров в вероятность ошибки. На основе выявленных закономерностей разработан и валидирован модуль адресной предобработки, нормализующий отклоняющиеся параметры к референсным диапазонам³. Полученные результаты позволили сформулировать обобщенный методологический подход, применимый не только к конкретным дерматологическим программам, но и к любым медицинским ПКЗ на основе ИИ, работающим с фотоизображениями.

Цель – разработка обобщенной методологии анализа причин ошибочной классификации фотоизображений ПКЗ на основе моделей ИИ, обеспечивающей установление референсных диапазонов фотометрических и текстурных параметров для конкретной области медицинской визуализации, а также формирование критериев исключения изображений с аномальными значениями параметров при работе модулей предобработки в составе медицинских приложений.

МАТЕРИАЛ И МЕТОДЫ / MATERIAL AND METHODS

Общая структура методологии / Methodology general structure

Методология анализа причин ошибочной классификации фотоизображений ПКЗ организована как последовательность из 8 связанных этапов, обеспечивающих движение от исходного набора фотоизображений к практическому результату в виде формализованных референсных диапазонов параметров и встроенного в приложение модуля исключения фотоизображений с аномальными значениями метрик. Концептуальная схема методологии представлена на **рисунке 1**.

³ Свидетельство о государственной регистрации программы для ЭВМ № 2026616231 Российская Федерация. Программа для автоматизированной визуальной идентификации и дифференциальной диагностики доброкачественных и злокачественных опухолей кожи с модулем нормализации параметров фотоизображений и алгоритмами маршрутизации пациентов “Derma Onko Melanoma Check Plus” : заявл. 04.03.2026 : опубл. 04.03.2026 / Д.И. Корабельников, А.И. Ламоткин ; заявитель Автономная некоммерческая организация дополнительного профессионального образования «Московский медико-социальный институт имени Ф.П. Гааза».

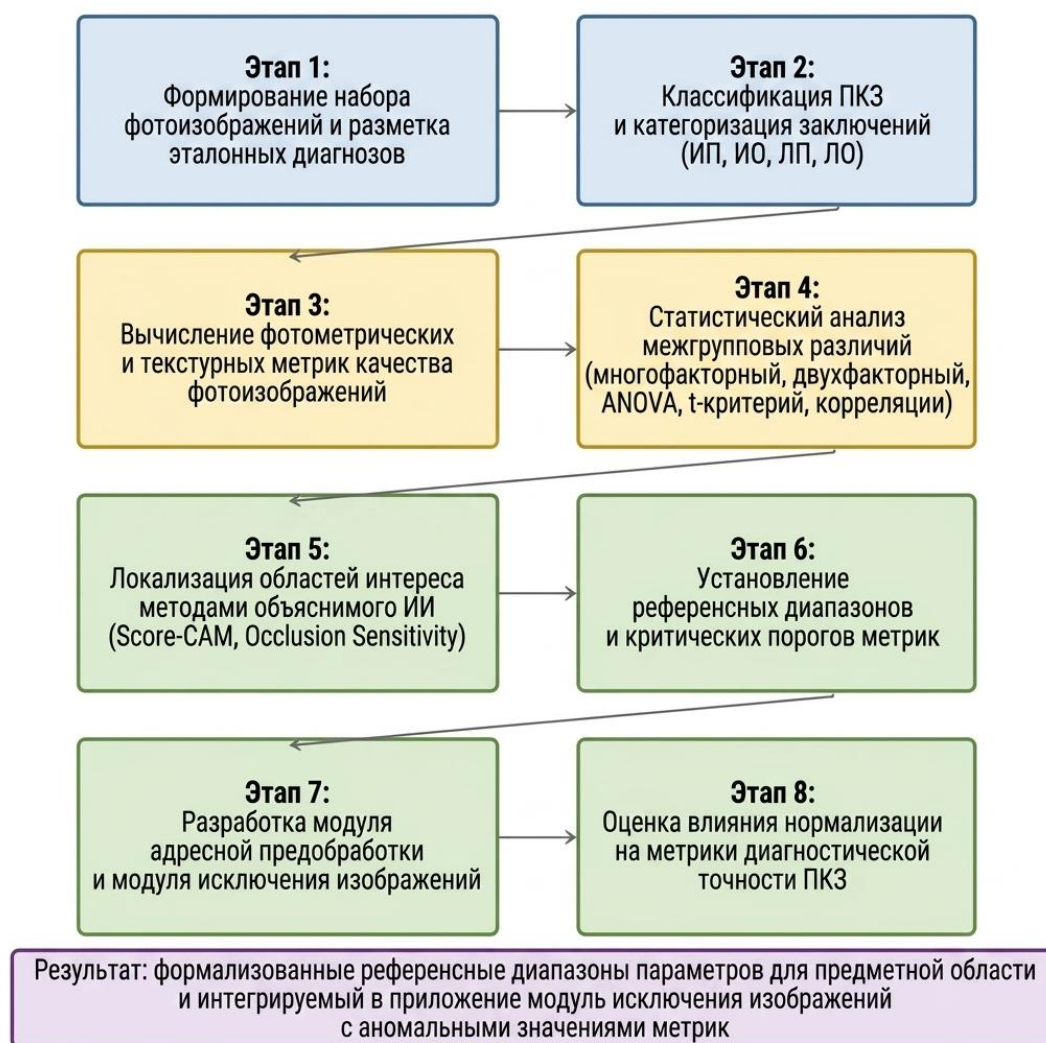


Рисунок 1. Концептуальная схема методологии анализа причин ошибочной классификации фотоизображений программами компьютерного зрения (ПКЗ) на основе искусственного интеллекта (ИИ).

ИП – истинно положительное; ИО – истинно отрицательное; ЛП – ложноположительное; ЛО – ложноотрицательное; ANOVA (англ. ANalysis Of VAriance) – дисперсионный анализ; Score-CAM (англ. score-weighted class activation mapping) – метод картирования активации классов, в котором для взвешивания карт признаков используются не градиенты, а значения уверенности модели; Occlusion Sensitivity – чувствительность к окклюзии, метод анализа, который помогает понять, какие части изображения или данных наиболее важны для классификации глубокой нейронной сети

Figure 1. Conceptual framework of the methodology for analyzing the causes of erroneous classification of photographic images by artificial intelligence (AI)-driven computer vision programs (CVPs).

TP – true positive; TN – true negative; FP – false positive; FN – false negative; ANOVA (ANalysis Of VAriance) – a statistical method used to assess differences between group means;

Score-CAM (score-weighted class activation mapping) – a class activation mapping technique that weights feature maps using the model’s confidence scores rather than gradients; Occlusion Sensitivity – an interpretability method used to identify which regions of an image or input data are most important for the classification decisions of a deep neural network

Этапы 1 и 2 формируют выборку фотоизображений и структурируют ее по четырем стандартным категориям заключений ПКЗ. Этапы 3, 4 и 5 раскрывают источники ошибок посредством количественного описания фотометрических и текстурных параметров, статистической верификации межгрупповых различий и локализации значимых для решения модели областей фотоизображения методами объяснимого ИИ. Этапы 6, 7 и 8 обеспечивают переход от установленных закономерностей к практическим инструментам: формализации референсных диапазонов, разработке модуля адресной предобработки и количественной оценке его влияния на метрики диагностической точности ПКЗ (табл. 1).

Таблица 1. Этапы методологии анализа причин ошибочной классификации фотоизображений программами компьютерного зрения (ПКЗ)

Table 1. Steps of the methodology for analyzing the causes of erroneous classification of photo images by computer vision programs (CVPs)

| № / No. | Наименование этапа / Step | Содержание / Content | Результат / Result |
|---------|---|--|--|
| 1 | Формирование набора фотоизображений / Formation of a dataset of photographic images | Сбор фотоизображений с верифицированными эталонными диагнозами, стандартизация метаданных / Collection of photographic images with verified reference diagnoses, standardization of metadata | Анонимизированная база данных / Anonymized database |
| 2 | Классификация ПКЗ и категоризация заключений / Classification of CVPs and categorization of conclusions | Инференс через модель ПКЗ, распределение по категориям ИП, ИО, ЛП, ЛО / Inference through the CVP model, distribution by categories of TP, TN, FP, FN | Размеченный по категориям набор / Set marked by categories |
| 3 | Вычисление метрик качества / Quality metrics calculation | Расчет фотометрических и текстурных параметров каждого фотоизображения / Calculation of photometric and textural metrics of each image | Матрица параметров изображений / Image parameter matrix |

| | | | |
|---|---|--|--|
| 4 | Статистический анализ / Statistical analysis | Однофакторный и двухфакторный дисперсионный анализ ANOVA, критерий Уэлча, корреляционный анализ по Спирмену, многомерный дисперсионный анализ MANOVA, множественная логистическая регрессия с контролем VIF / One-way and two-way ANOVA, Welch's test, Spearman's correlation analysis, multivariate MANOVA, multiple logistic regression with VIF control | Карта статистически значимых различий / Map of statistically significant differences |
| 5 | Локализация областей интереса / Localization of regions of interest | Применение методов объяснимого ИИ (Score-CAM, Occlusion Sensitivity) / Application of explainable AI methods (Score-CAM, Occlusion Sensitivity) | Тепловые карты значимости / Importance heat maps |
| 6 | Установление референсных диапазонов / Establishing reference ranges | Расчет интервалов [mean – 2·std; mean + 2·std] по категориям ИП и ИО / Calculation of intervals [mean – 2·std; mean + 2·std] for IP and IO categories | Нормативные значения параметров / Standard values of parameters |
| 7 | Разработка модуля адресной предобработки / Development of targeted preprocessing module | Реализация последовательной коррекции отклоняющихся параметров / Implementation of sequential correction of deviating parameters | Программный модуль на языке Python / Python program module |
| 8 | Оценка влияния нормализации / Impact of normalization assessment | Повторный инференс предобработанных фотоизображений; расчет метрик точности / Repeated inference of preprocessed photographic images; calculation of accuracy metrics | Прирост точности, чувствительности, специфичности / Improvement in accuracy, sensitivity, and specificity |

Примечание. ИП – истинно положительное; ИО – истинно отрицательное; ЛП – ложноположительное; ЛО – ложноотрицательное; ANOVA (англ. ANalysis Of VAriance) – дисперсионный анализ; MANOVA (англ. Multivariate ANalysis Of VAriance) – многомерный дисперсионный анализ; VIF (англ. variance inflation factor) – фактор инфляции дисперсии; ИИ – искусственный интелект; mean – среднее значение; std (англ. standard deviation) – стандартное отклонение; Score-CAM (англ. class activation mapping) – метод картирования активации классов, в котором для взвешивания карт признаков используются не градиенты, а значения уверенности модели; Occlusion Sensitivity – чувствительность к окклюзии, метод анализа, который помогает понять, какие части изображения или данных наиболее важны для классификации глубокой нейронной сети.

Note. TP – true positive; TN – true negative; FP – false positive; FN – false negative; ANOVA – analysis of variance; MANOVA – multivariate analysis of variance; VIF – variance inflation factor; AI – artificial

intelligence; mean – mean value; std – standard deviation; Score-CAM (score-weighted class activation mapping) – a class activation mapping method that uses model confidence scores rather than gradients to weight feature maps; Occlusion Sensitivity – an analysis technique used to determine which regions of an image or input data are most important for the classification decisions of a deep neural network.

Формирование набора фотоизображений и эталонных диагнозов / Forming a set of photo images and reference diagnoses

Каждое фотоизображение, включаемое в анализ, должно сопровождаться верифицированным эталонным диагнозом, установленным независимым от модели ИИ способом. В дерматологии таким способом служат гистологическое исследование биопсийного материала и заключение врача-специалиста по дерматоскопии, в офтальмологии – заключение врача-офтальмолога по результатам комплексного обследования с применением оптической когерентной томографии, в гистологии – заключение врача-патологоанатома по микропрепарату.

Стандартизация метаданных является обязательным условием воспроизводимости анализа. Для каждого фотоизображения регистрируются идентификатор пациента и фотоизображения, пол, возраст, верифицированный диагноз по Международной классификации болезней 10-го пересмотра, анатомическая локализация, тип фотоизображения, дата формирования, модель регистрирующего устройства, разрешение, формат файла и MD5-хэш⁴ (англ. Message Digest – сводка сообщений и англ. hash function от hash — превращать в фарш, мелкие кусочки, или, в данном случае, функция свёртки — функция, преобразующая массив входных данных произвольного размера в выходную битовую строку установленного размера в соответствии с определённым алгоритмом) - результат преобразования массива входных данных в выходную битовую строку установленного размера в соответствии с определённым алгоритмом. Подобный состав метаданных позволяет в дальнейшем стратифицировать анализ по подгруппам и контролировать варибельность, обусловленную техническими характеристиками регистрирующих устройств.

Категоризация заключений ПКЗ / CVP output categorization

Сопоставление заключения ПКЗ с эталонным диагнозом формирует четыре стандартные категории, образующие диагностическое поле классификации (**рис. 2**):

⁴ MD5-хэш (англ. message digest – сводка сообщений; англ. hash – в данном случае функция свертки) – функция, преобразующая массив входных данных произвольного размера в выходную битовую строку установленного размера в соответствии с определённым алгоритмом.

- истинно положительные (ИП) – случаи, когда положительный эталонный диагноз верно распознан моделью ИИ;
- истинно отрицательные (ИО) – случаи, когда отрицательный эталонный диагноз верно распознан моделью ИИ;
- ложноположительные (ЛП) – случаи, когда отрицательный эталонный диагноз ошибочно классифицирован моделью как положительный;
- ложноотрицательные (ЛО) – случаи, когда положительный эталонный диагноз ошибочно классифицирован моделью как отрицательный.



Рисунок 2. Четырехпольная категоризация заключений программ компьютерного зрения (ПКЗ) по соотношению эталонного диагноза и заключения модели искусственного интеллекта.

ЗНО – злокачественное новообразование; ДНО – доброкачественное новообразование

Figure 2. Four-category classification of computer vision program (CVP) outputs based on the correspondence between reference diagnosis and artificial intelligence (AI) model output. MN – malignant neoplasm; BN – benign neoplasm

Категории ИП и ИО объединяются как корректные классификации и в дальнейшем используются для установления референсных диапазонов параметров. Категории ЛП и ЛО выступают объектом анализа причин ошибочной классификации.

Вычисление фотометрических и текстурных метрик качества / Calculating photometric and textural quality metrics

Для каждого фотоизображения в цветовой модели RGB (англ. red, green, blue – красный, зеленый, синий), описываемого с помощью uint8⁵, вычисляется набор из 13 количественных метрик, охватывающих ключевые аспекты фотометрического и текстурного качества. Перечень метрик и их интерпретация представлены в **таблице 2**.

Таблица 2. Перечень фотометрических и текстурных метрик качества фотоизображений

Table 2. List of photometric and textural metrics of photo image quality

| Метрика / Metric | Обозначение / Designation | Определение / Definition | Диагностическое значение / Diagnostic value |
|---|---------------------------|---|---|
| Яркость / Brightness | brightness | Среднее значение яркостного канала на нормализованной шкале 0–1 / Mean brightness channel value on a normalized 0–1 scale | Контроль экспозиции / Exposure control |
| Контраст / Contrast | contrast | Стандартное отклонение яркостного канала / Standard deviation of brightness channel | Информативность границ / Edge informativeness |
| Резкость / Sharpness | sharpness | Дисперсия результата применения оператора Лапласа / Variance of the Laplacian operator response | Контроль фокусировки / Focus control |
| Энтропия / Entropy | entropy | Информационная энтропия гистограммы яркости / Information entropy of brightness histogram | Сложность текстуры / Texture complexity |
| Высокочастотная насыщенность по FFT / High-frequency content (FFT blur) | fft_blur | Среднее значение логарифмической амплитуды двумерного FFT / Mean logarithmic amplitude of the two-dimensional FFT | Контроль насыщенности/размытости // High-frequency content/blur control |

⁵ Тип данных, который используется для хранения 8-битных целых чисел без знака, – стандартный способ отображения изображений в модели RGB, где каждый пиксель описывается тройкой чисел в диапазоне от 0 до 255, соответствующих яркостям базовых составляющих.

| | | | |
|--|--------------|---|---|
| Доля пересвеченных пикселей / Percentage of overexposed pixels | clipped_high | Доля пикселей с $Y > 0,95$ / Percentage of pixels with $Y > 0.95$ | Контроль переэкспозиции / Overexposure control |
| Доля затемненных пикселей / Percentage of underexposed pixels | clipped_low | Доля пикселей с $Y < 0,05$ / Percentage of pixels with $Y < 0.05$ | Контроль недоэкспозиции / Underexposure control |
| Средние значения цветовых каналов // Mean values of color channels | | | |
| красный / red | mean_r | Среднее значение канала R на шкале 0–1 / Mean value of R channel on a 0–1 scale | Цветовая характеристика / Color characteristics |
| зеленый / green | mean_g | Среднее значение канала G на шкале 0–1 / Mean value of G channel on a 0–1 scale | |
| синий / blue | mean_b | Среднее значение канала B на шкале 0–1 / Mean value of B channel on a 0–1 scale | |
| Стандартные отклонения цветовых каналов // Standard deviations of color channels | | | |
| красный / red | std_r | Вариабельность канала R / R- channel variability | Однородность освещения / Illumination uniformity |
| зеленый / green | std_g | Вариабельность канала G / G- channel variability | |
| синий / blue | std_b | Вариабельность канала B / B- channel variability | |

Примечание. FFT (англ. fast Fourier transform) – быстрое преобразование Фурье.

Состав метрик подобран таким образом, чтобы охватить три самостоятельных измерения качества фотоизображения: фотометрическое (яркость, контраст, доли пересвеченных и затемненных пикселей), текстурно-информационное (энтропия, резкость, высокочастотная насыщенность) и цветовое (средние значения и стандартные отклонения каналов R, G, B). Подобная структура позволяет независимо идентифицировать источник отклонения и применять адресную коррекцию на этапе предобработки.

Статистический анализ межгрупповых различий / Statistical analysis of intergroup differences

Статистическая обработка матрицы параметров выполняется в программной среде Python с использованием библиотек Seaborn, Pandas, SciPy, Statsmodels⁶.

Для выявления систематических различий метрик качества фотоизображений между категориями заключений применяется однофакторный дисперсионный анализ ANOVA с последующей верификацией попарных различий по Т-критерий Уэлча (t-критерий неравных дисперсий). Корреляционный анализ по Спирмену позволяет оценить взаимозависимость метрик и обоснованно выбрать стратегию комплексной нормализации.

Визуализация распределений осуществляется посредством диаграмм типа «ящик с усами» (англ. boxplot), отражающих положение медианы, межквартильный размах и выбросы каждой метрики в разрезе категорий ИП, ИО, ЛП, ЛО. Парные диаграммы рассеяния (англ. pairplot) обеспечивают наглядное представление о корреляционной структуре метрик. Карты пространственного распределения значений пикселей дополняют статистический анализ сведениями о локализации различий между категориями на уровне отдельных областей фотоизображения.

Однофакторный дисперсионный анализ ANOVA устанавливает факт значимости различий метрики между категориями заключений, однако не позволяет ни оценить совместное действие нескольких факторов, ни количественно ранжировать вклад отдельных метрик в вероятность ошибочной классификации. Для содержательного раскрытия этих аналитических задач применяется многомерный статистический анализ, реализуемый тремя последовательно дополняющими друг друга методами.

Двухфакторный дисперсионный анализ ANOVA оценивает одновременное влияние двух факторов на каждую метрику качества. В качестве первого фактора выступает категория заключения (ИП, ИО, ЛП, ЛО), в качестве второго – применяемая программа ПКЗ. Дополнительно вычисляется эффект взаимодействия факторов, показывающий, остаются ли межгрупповые различия по категории заключения одинаковыми для разных программ. Незначимый эффект взаимодействия ($p \geq 0,05$) свидетельствует о воспроизводимости закономерностей и обосновывает применимость единого набора референсных диапазонов для всех исследованных программ ПКЗ.

Многомерный дисперсионный анализ MANOVA обобщает дисперсионный анализ на случай, когда зависимой переменной выступает не отдельная метрика, а совокупный

⁶ Seaborn – библиотека для создания статистических графиков, построенная на основе Matplotlib; Pandas – библиотека для работы с структурированными данными, преимущественно табличными (панельными); SciPy – библиотека для научных и инженерных расчетов, основанная на NumPy; Statsmodels – библиотека для статистического моделирования, проверки гипотез и анализа данных.

вектор метрик. Метод вычисляет четыре эквивалентных показателя многомерной статистики: лямбда Уилкса, след Пиллаи, след Хотеллинга–Лоули и наибольший корень Роя. Применение MANOVA принципиально оправданно в случае выраженной корреляционной структуры метрик, когда оценка различий по каждой метрике в отдельности приводит к избыточной множественности тестов и не учитывает совместного действия коррелирующих признаков.

Множественная логистическая регрессия позволяет количественно оценить вклад каждой метрики в вероятность ошибочного заключения с поправкой на остальные параметры. Зависимая переменная принимает два значения (1 – ошибочное заключение категорий ЛП или ЛО; 0 – правильное заключение категорий ИП или ИО), независимыми переменными выступают все 13 метрик качества после z-стандартизации, обеспечивающей сопоставимость коэффициентов. Отношения шансов (англ. odds ratio, OR), полученные экспоненцированием коэффициентов, характеризуют, во сколько раз изменяется вероятность ошибки при изменении соответствующего предиктора на одно стандартное отклонение. Контроль мультиколлинеарности осуществляется по фактору инфляции дисперсии (англ. variance inflation factor, VIF) с пороговым значением 10; предикторы с VIF выше порога исключаются из окончательной модели для предотвращения нестабильности оценок коэффициентов.

Принципиально важно, что MANOVA применяется не вместо однофакторного ANOVA, а в дополнение к нему. Совпадение значимых факторов в обеих моделях служит независимым подтверждением воспроизводимости установленных закономерностей. Новые результаты, доступные только при многофакторной постановке (эффект взаимодействия, показатели многомерной статистики, OR), уточняют интерпретацию и обеспечивают ее количественную ясность.

Локализация областей интереса методами объяснимого ИИ / Localizing regions of interest using explainable AI methods

Локализация областей фотоизображений, определяющих решение нейронной сети, осуществляется двумя дополняющими друг друга методами. Метод взвешенного картирования активации классов (Score-CAM) показывает, на какие участки фотоизображения опирается глубокая нейронная сеть при принятии решения, посредством построения тепловых карт значимости. Метод определения чувствительности к окклюзии (Occlusion Sensitivity) последовательно закрывает небольшие участки фотоизображения и измеряет влияние такой окклюзии на уверенность модели в предсказании класса.

Совместное применение двух методов обеспечивает более надежную локализацию значимых областей: Score-CAM выявляет области положительного вклада в решение модели, а Occlusion Sensitivity дополнительно различает области, окклюзия которых снижает уверенность модели в правильном классе, и области, окклюзия которых, напротив, повышает ее. Несовпадение значимых областей с клинически информативными зонами фотоизображения служит непосредственным индикатором ошибочной логики модели.

Установление референсных диапазонов и критических порогов / Establishing reference ranges and critical thresholds

Референсные диапазоны параметров фотоизображений устанавливаются по характеристикам правильно классифицированных фотоизображений (объединенная выборка категорий ИП и ИО). Для каждой метрики рассчитывается интервал [$\text{mean} - 2 \cdot \text{std}$; $\text{mean} + 2 \cdot \text{std}$], охватывающий примерно 95% правильно классифицированных случаев в предположении приближенно нормального распределения. Полученные интервалы рассматриваются как нормативные значения параметра для конкретной предметной области и конкретной модели ПКЗ.

Принципиальное значение имеет тот факт, что референсные диапазоны являются характеристикой не только модели, но и предметной области. В дерматологии при макросъемке кожных покровов нормативный диапазон яркости составляет 0,467–0,942 для меланоцитарных опухолей и 0,548–0,880 для немеланоцитарных, что отражает различия в пигментной архитектуре и условиях съемки. В офтальмологии нормативный диапазон яркости при фундус-фотографировании смещен в сторону более низких значений в связи с особенностями оптической схемы фундус-камер и пигментацией глазного дна. В гистологии нормативные значения яркости и контраста определяются протоколом окрашивания микропрепаратов и параметрами сканирующей оптики, однако в практике обучения моделей ИИ для анализа цифровых гистологических препаратов общепринятым этапом является нормализация по окрашиванию (англ. stain normalization), приводящая интенсивность и распределение цветовых каналов сканированных препаратов к референсному образцу.

В связи с этим систематические отклонения яркости или контраста, выявляемые в гистологических наборах данных, как правило, не обусловлены условиями получения фотоизображения, а указывают на некорректно выбранную методологию подготовки обучающей выборки, в частности на отсутствие или недостаточную применимость алгоритмов нормализации окрашивания на этапе обучения. Подобная закономерность

принципиально отличает гистологию от дерматологии и офтальмологии, где источником отклонений выступают преимущественно условия фотосъемки, и должна учитываться при интерпретации результатов анализа ошибочных заключений в каждой области. Сопоставление условий получения фотоизображений в различных областях медицинской визуализации представлено в **таблице 3**.

Таблица 3. Сопоставление условий получения фотоизображений в различных областях медицинской визуализации

Table 3. Comparison of photo image acquisition conditions across different medical imaging domains

| Область / Domain | Регистрирующее устройство / Imaging device | Источник освещения / Illumination source | Особенности референса / Reference specifics |
|---|--|--|--|
| Дерматология / Dermatology | Камера смартфона, цифровой фотоаппарат с макрообъективом / Smartphone camera, digital camera with a macro lens | Рассеянный видимый свет, LED-вспышка / Diffuse visible light, LED flash | Высокая вариабельность яркости и цветовых каналов / High variability in brightness and color-channel values |
| Офтальмология (фундус) / Ophthalmology (fundus imaging) | Специализированная фундус-камера / Dedicated fundus camera | Инфракрасная подсветка, ксеноновая вспышка / Infrared illumination, xenon flash | Смещение красного и зеленого каналов, низкая средняя яркость / Shifted red and green channel distributions, low mean brightness |
| Гистология / Histology | Сканер микропрепаратов / Whole-slide scanner | Стандартизованная подсветка препарата / Standardized specimen illumination | Жестко регламентированные значения параметров, обязательная нормализация по окрашиванию / Strictly regulated parameter values; mandatory stain normalization |
| Эндоскопия / Endoscopy | Эндоскопическая видеокамера / Endoscopic video camera | Холодный свет через световод / Cold light delivered through a light guide | Бликовость, искажения цветопередачи / Specular reflections and color-rendering distortions |
| УЗИ-диагностика / US imaging | УЗИ-аппарат с цифровым выходом / US system with digital output | Внутренняя визуализация акустического сигнала / Internal visualization of acoustic signals | Черно-белые изображения, высокая шумность / Grayscale images with high noise levels |

Примечание. LED (англ. light-emitting diode) – светодиод; УЗИ – ультразвуковое исследование.

Note. LED – light-emitting diode; US – ultrasound.

Критические пороги задают границы пригодности фотоизображения для инференса. Критический порог 0,55 установлен эмпирическим способом по 99-му процентилю доли пересвеченных и затемнённых пикселей в выборке правильно классифицированных фотоизображений. При превышении доли пересвеченных пикселей значения 0,55 (более 55% площади фотоизображения) восстановление диагностически значимой информации становится принципиально невозможным независимо от качества алгоритмов предобработки.

Аналогичное правило действует для доли затемненных пикселей и для крайних значений общей яркости и контраста. Фотоизображения, у которых хотя бы одна метрика превышает критический порог, исключаются из инференса с уведомлением пользователя о необходимости повторного фотографирования.

Разработка модуля адресной предобработки / Developing address preprocessing module

Модуль адресной предобработки реализует принцип метрически обоснованной нормализации: каждое фотоизображение проходит многоэтапную коррекцию, при которой каждый этап активируется только при обнаружении значимого отклонения соответствующего параметра от установленного референсного диапазона. Пороговое значение отклонения (англ. deviation threshold) принимается равным 5% от ширины нормального диапазона ($\text{deviation_threshold} = 0,05$). Алгоритм работы модуля представлен на **рисунке 3**.

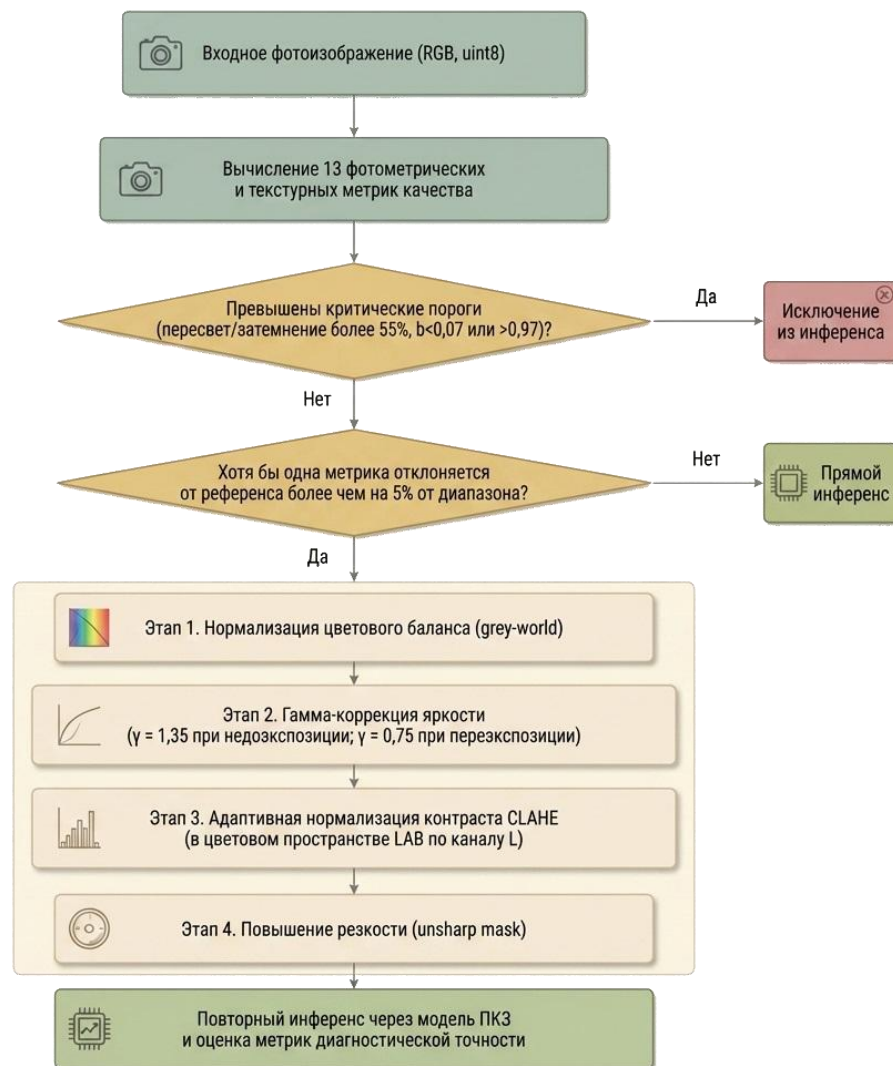


Рисунок 3. Алгоритм работы модуля адресной предобработки фотоизображений: проверка критических порогов, контроль отклонения метрик от референсного диапазона и последовательное применение четырех этапов коррекции.

CLAHE (англ. Contrast Limited Adaptive Histogram Equalization) – алгоритм обработки изображений, который улучшает локальный контраст, но ограничивает усиление шума в относительно гомогенных областях; LAB – сокращенное название цветового пространства CIE 1976 L a b*; ПКЗ – программа компьютерного зрения

Figure 3. Algorithm of targeted photo image preprocessing module: verification of critical thresholds, assessment of metric deviations from the reference range, and sequential application of four correction stages.

CLAHE (Contrast Limited Adaptive Histogram Equalization) – an image-processing algorithm that enhances local contrast while limiting noise amplification in relatively homogeneous regions; LAB – abbreviated designation of the CIE 1976 L*a*b* color space; CVP – computer vision program

На этапе 1 при отклонении любого из средних значений цветовых каналов выполняется автоматический баланс белого методом серого мира (англ. grey-world). Метод вычисляет среднее по каждому каналу, находит их глобальное среднее и масштабирует каждый канал так, чтобы его среднее совпало с глобальным средним, что устраняет систематические цветовые смещения, обусловленные условиями освещения.

На этапе 2 при отклонении яркости ниже нормального диапазона применяется повышающая гамма-коррекция ($\gamma=1,35$), при отклонении выше диапазона – понижающая ($\gamma=0,75$). Коррекция реализуется через предварительно вычисленную таблицу преобразования цветов (англ. look-up table, LUT), что обеспечивает ее вычислительную эффективность.

На этапе 3 при отклонении контраста или энтропии применяется алгоритм адаптивной нормализации контраста с ограничением (англ. Contrast Limited Adaptive Histogram Equalization, CLAHE) в цветовом пространстве LAB (CIE 1976 L a b). Коррекция выполняется только по яркостному каналу L без изменения цветовых каналов A и B, что сохраняет естественность цветопередачи. Параметр ограничения контраста (англ. clipLimit) и размер плитки настраиваются с учетом специфики предметной области.

На четвертом этапе при значении резкости ниже нижней границы нормального диапазона применяется маска нерезкости (англ. unsharp mask). Метод вычисляет разность между оригинальным фотоизображением и его гауссовским размытием и прибавляет взвешенную разность к оригиналу, усиливая высокочастотные компоненты, критически важные для распознавания тонких текстурных особенностей.

Оценка эффективности нормализации / Evaluating normalization effectiveness

Оценка влияния модуля адресной предобработки на работу ПКЗ осуществляется путем сравнения метрик диагностической точности до и после нормализации. Используются три классические метрики: точность (англ. accuracy, Ac), чувствительность (англ. sensitivity, Se) и специфичность (англ. specificity, Sp), которые рассчитываются по формулам:

$$Ac = (\text{ИП} + \text{ИО}) / (\text{ИП} + \text{ИО} + \text{ЛП} + \text{ЛО}),$$

$$Se = \text{ИП} / (\text{ИП} + \text{ЛО}),$$

$$Sp = \text{ИО} / (\text{ИО} + \text{ЛП}).$$

Сравнение проводится в одном и том же наборе фотоизображений, что обеспечивает корректность вывода о причинной связи между нормализацией параметров и приростом метрик.

РЕЗУЛЬТАТЫ / RESULTS

Распределение метрик качества по категориям заключений / Distributing quality metrics by output categories

Применение методологии на наборе из 460 фотоизображений меланокитарных опухолей кожи обнаружило систематические межгрупповые различия фотометрических и текстурных метрик. ИП результаты характеризовались пониженной яркостью (медиана 0,638 на нормализованной шкале 0–1) и повышенной энтропией (медиана 5,159), свидетельствующей о сложной текстуре с клинически информативными деталями. ЛП результаты, напротив, демонстрировали повышенную яркость (медиана 0,740) и сниженную энтропию (медиана 4,898), что соответствует переэкспонированным фотоизображениям с размытой текстурной детализацией. Распределение основных метрик по категориям представлено на **рисунке 4**.

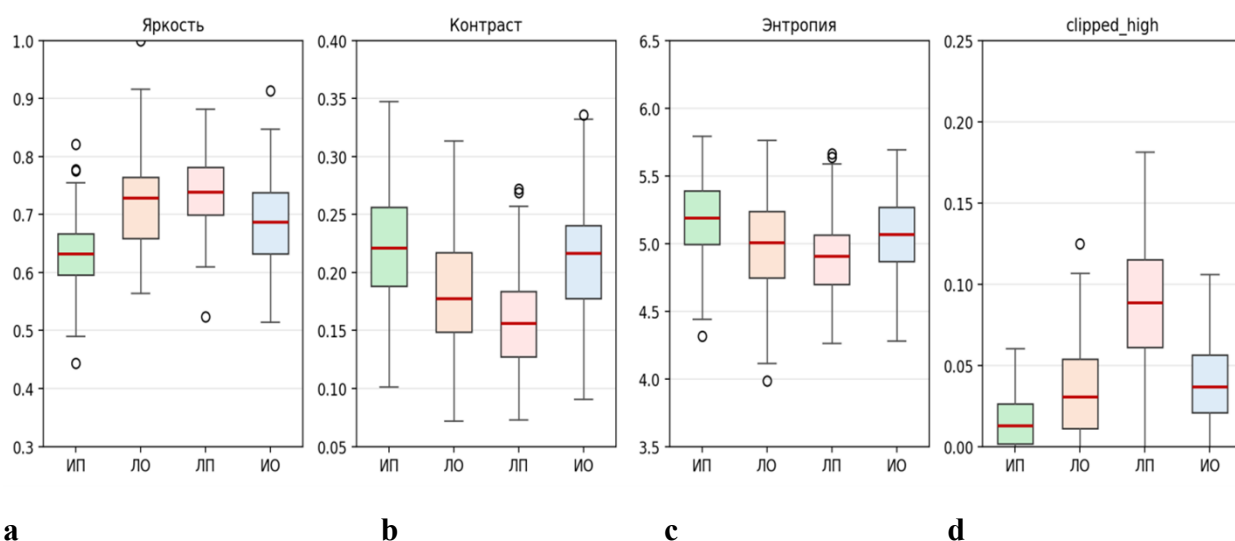


Рисунок 4. Распределение фотометрических и текстурных метрик качества фотоизображений по категориям заключений программ компьютерного зрения:

a – яркость; **b** – контраст; **c** – энтропия; **d** – доля пересвеченных пикселей.

ИП – истинно положительное; ЛО – ложноотрицательное; ЛП – ложноположительное; ИО – истинно отрицательное

Figure 4. Distribution of photometric and textural quality metrics of photo images across computer vision program outcome categories:

a – brightness; **b** – contrast; **c** – entropy; **d** – percentage of overexposed pixels.

TP – true positive; FN – false negative; FP – false positive; TN – true negative

По данным однофакторного ANOVA межгрупповые различия достигли уровня статистической значимости для яркости ($F=14,83-23,36$; $p<0,001$), контраста ($F=18,97-39,31$; $p<0,001$), энтропии ($F=13,50-18,34$; $p<0,001$), средних значений красного ($F=16,05-24,47$; $p<0,001$) и зеленого ($F=14,53-22,77$; $p<0,001$) каналов. Критерий Уэлча верифицировал воспроизводимость попарных различий между категориями: наиболее выраженные расхождения зафиксированы между ИО и ИП заключениями по яркости ($t=7,44-9,10$; $p<0,001$), энтропии ($t=-8,23...-10,40$; $p<0,001$) и контрасту ($t=-7,39...-9,85$; $p<0,001$), а также между ИП и ЛП заключениями по яркости ($t=-4,11...-6,71$; $p<0,001$) и средним значениям цветовых каналов ($t=3,35-4,32$; $p<0,01$).

Результаты многомерного анализа / Results of multivariate analysis

Двухфакторный ANOVA подтвердил статистически значимое влияние категории заключения по всем 13 исследованным метрикам ($p<0,001$ для одиннадцати метрик, $p<0,01$ для метрик резкости и доли пересвеченных пикселей). Наиболее выраженные эффекты зафиксированы для стандартного отклонения красного канала ($F=72,29$), контраста ($F=54,39$), средних значений красного ($F=46,63$) и зеленого ($F=40,34$) каналов, а также для яркости ($F=41,49$) и стандартного отклонения зеленого канала ($F=45,64$). Все эти метрики совпадают со значимыми по результатам однофакторного ANOVA, что свидетельствует о воспроизводимости установленных закономерностей. Сводные результаты представлены в **таблице 4**.

Таблица 4. Результаты двухфакторного дисперсионного анализа ANOVA метрик качества фотоизображений по факторам «категория заключения» и «программа компьютерного зрения»

Table 4. Results of the two-way ANOVA of photo image quality metrics by the factors “output category” and “computer vision program”

| Метрика / Metric | Категория заключения / Output category | | ПКЗ / CVP | | Взаимодействие / Interaction | |
|----------------------|--|--------|-----------|-------|------------------------------|-------|
| | F | p | F | p | F | p |
| Яркость / Brightness | 41,49 | <0,001 | 0,04 | 0,844 | 0,79 | 0,502 |
| Контраст / Contrast | 54,39 | <0,001 | 0,05 | 0,822 | 0,52 | 0,668 |
| Резкость / Sharpness | 5,70 | <0,001 | 0,02 | 0,888 | 0,28 | 0,838 |

| | | | | | | |
|--|-------|--------|------|-------|------|-------|
| Энтропия / Entropy | 21,21 | <0,001 | 0,35 | 0,554 | 1,81 | 0,145 |
| Высокочастотная насыщенность по FFT / High-frequency content (FFT blur) | 7,36 | <0,001 | 0,13 | 0,721 | 0,59 | 0,623 |
| Доля пересвеченных пикселей / Percentage of overexposed pixels | 5,41 | 0,001 | 0,59 | 0,444 | 0,79 | 0,501 |
| Доля затемненных пикселей / Percentage of underexposed pixels | 14,36 | <0,001 | 0,00 | 0,962 | 1,21 | 0,306 |
| Средние значения цветовых каналов / Mean values for color channels | | | | | | |
| mean_r | 46,63 | <0,001 | 0,16 | 0,690 | 1,00 | 0,393 |
| mean_g | 40,34 | <0,001 | 0,01 | 0,934 | 0,56 | 0,639 |
| mean_b | 20,23 | <0,001 | 0,00 | 0,965 | 0,16 | 0,924 |
| Стандартные отклонения цветовых каналов / Standard deviations for color channels | | | | | | |
| std_r | 72,29 | <0,001 | 0,15 | 0,701 | 1,07 | 0,361 |
| std_g | 45,64 | <0,001 | 0,24 | 0,626 | 0,34 | 0,796 |
| std_b | 17,43 | <0,001 | 0,74 | 0,390 | 0,31 | 0,821 |

Примечание. ПКЗ – программа компьютерного зрения; FFT (англ. fast Fourier transform) – быстрое преобразование Фурье.

Note. CVP – computer vision program; FFT – fast Fourier transform.

Принципиально новый результат двухфакторной модели – незначимость фактора ПКЗ во всех 13 метриках ($p=0,39-0,96$) и эффекта взаимодействия категорий заключения и ПКЗ ($p=0,15-0,92$). Подобный результат свидетельствует о том, что входные распределения параметров фотоизображений эквивалентны для обеих исследованных программ, а направление и величина межгрупповых различий по категории заключения воспроизводятся независимо от выбора программы. Это обосновывает применимость единого набора референсных диапазонов на этапе адресной предобработки.

Анализ MANOVA подтвердил установленные закономерности на уровне совокупного вектора фотометрических и текстурных метрик. Для фактора категории заключения значение лямбды Уилкса составило 0,639 ($F=10,37$; $df=21/1289,8$; $p<0,001$), следа Пиллаи – 0,380 ($F=9,35$; $p<0,001$), следа Хотеллинга-Лоули – 0,535 ($F=11,41$; $p<0,001$). Для фактора ПКЗ показатели многомерной статистики оказались статистически незначимыми (лямбда Уилкса 0,998; $F=0,138$; $p=0,995$), что согласуется с результатами двухфакторного ANOVA по отдельным метрикам и подтверждает независимость закономерностей от выбора программы.

Множественная логистическая регрессия позволила количественно ранжировать вклад отдельных метрик в вероятность ошибочного заключения с поправкой на остальные параметры. Контроль мультиколлинеарности выявил высокие значения VIF (10,2–931,1) для метрик яркости, контраста, средних значений и стандартных отклонений цветовых каналов, что обусловлено установленными корреляционными связями между этими признаками. Указанные метрики были исключены из логистической модели; в окончательную модель вошли 5 независимых предикторов: резкость, энтропия, высокочастотная насыщенность по быстрому преобразованию Фурье (англ. fast Fourier transform, FFT), доля пересвеченных и доля затемненных пикселей. Качество модели характеризуется псевдокоэффициентом детерминации МакФаддена (англ. McFadden's pseudo R-squared) 0,085 при статистически значимом отношении правдоподобия (англ. log-likelihood ratio, LLR) ($p < 0,001$). Результаты представлены в **таблице 5**.

Таблица 5. Результаты множественной логистической регрессии: количественный вклад метрик качества фотоизображений в вероятность ошибочного заключения

Table 5. Results of multiple logistic regression: quantitative contribution of photo image quality metrics to the probability of an erroneous conclusion

| Предиктор / Predictor | Коэффициент / Coefficient | z | p | OR | 95% CI |
|---|---------------------------|-------|--------|------|-----------|
| Высокочастотная насыщенность по FFT / High-frequency content (FFT blur) | +1,1245 | 3,97 | <0,001 | 3,08 | 1,77–5,36 |
| Резкость / Sharpness | -1,1818 | -3,07 | 0,002 | 0,31 | 0,14–0,65 |
| Доля пересвеченных пикселей / Percentage of overexposed pixels | +0,4918 | 2,50 | 0,013 | 1,64 | 1,11–2,41 |
| Энтропия / Entropy | +0,2914 | 1,37 | 0,170 | 1,34 | 0,88–2,03 |
| Доля затемненных пикселей / Percentage of underexposed pixels | +0,2017 | 1,69 | 0,092 | 1,22 | 0,97–1,55 |

Примечание. FFT (англ. fast Fourier transform) – быстрое преобразование Фурье; OR (англ. odds ratio) – отношение шансов; CI (англ. confidence interval) – доверительный интервал.

Note. FFT – fast Fourier transform; OR – odds ratio; CI – confidence interval.

Статистически значимыми предикторами вероятности ошибочного заключения оказались высокочастотная насыщенность (размытие) по FFT, резкость и доля пересвеченных пикселей. Полученные значения отношений шансов для z-стандартизованных предикторов позволяют интерпретировать эффекты следующим образом:

– уменьшение высокочастотной насыщенности по FFT - увеличение размытия на одно стандартное отклонение повышает шансы ошибочного заключения в 3,08 раза (на 208%);

– увеличение резкости на одно стандартное отклонение, напротив, снижает шансы ошибочного заключения на 69% (отношение шансов 0,31);

– увеличение доли пересвеченных пикселей на одно стандартное отклонение повышает шансы ошибки в 1,64 раза (на 64%).

Полученные результаты не противоречат, а уточняют выводы однофакторного ANOVA: наибольший количественный вклад в вероятность ошибочного заключения вносят показатели текстурной информативности и фокусировки фотоизображения.

Корреляционная структура метрик качества / Correlation structure of quality metrics

Парные диаграммы рассеяния и корреляционная матрица (рис. 5) выявили устойчивые зависимости между метриками. Яркость и энтропия находились в выраженной отрицательной связи ($r=-0,76$): фотоизображения с насыщенной текстурной детализацией, как правило, характеризовались умеренной яркостью. Контраст положительно коррелировал с энтропией ($r=0,56$), что закономерно, поскольку высококонтрастные фотоизображения воспроизводят тонкие детали клинически значимого рисунка. Резкость демонстрировала умеренную положительную связь с энтропией ($r=0,26$) и умеренную отрицательную с яркостью ($r=-0,33$).

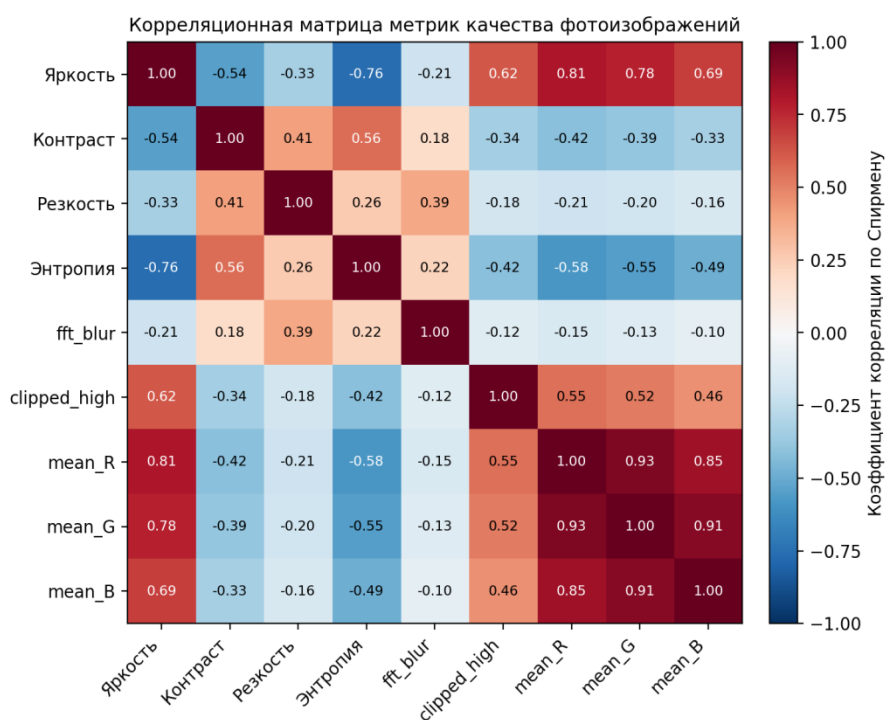


Рисунок 5. Корреляционная матрица фотометрических и текстурных метрик качества фотоизображений (коэффициент корреляции по Спирмену)

Figure 5. Correlation matrix of photometric and textural metrics of photo image quality (Spearman's rank correlation coefficient)

Выявленная отрицательная корреляция яркости и энтропии указывает на то, что переэкспонированные фотоизображения теряют именно те текстурные детали, которые служат ключевыми диагностическими признаками. Это означает, что модель ИИ затрудняется верно классифицировать фотоизображения с одновременно высокой яркостью и сниженной энтропией. Установленные взаимосвязи метрик объясняют высокие значения VIF, выявленные в логистической регрессии, и обосновывают примененную процедуру контроля.

Установленные референсные диапазоны и критические пороги / Established reference ranges and critical thresholds

На основе характеристик правильно классифицированных фотоизображений (объединенная выборка категорий ИП и ИО) рассчитаны референсные диапазоны метрик качества для двух самостоятельных дерматологических наборов: меланоцитарных и немеланоцитарных (эпидермальных) опухолей кожи. Полученные диапазоны представлены в **таблице 6** и подтверждают принципиальное положение методологии о необходимости установления референсных диапазонов отдельно для каждой предметной подобласти.

Таблица 6. Референсные диапазоны метрик качества фотоизображений для меланоцитарного и эпидермального датасетов (минимальные и максимальные значения)

Table 6. Reference ranges of photo image quality metrics for melanocytic and epidermal datasets (minimum and maximum values)

| Метрика / Metric | Датасет / Dataset | |
|---|------------------------------|---------------------------|
| | Меланоцитарный / Melanocytic | Эпидермальный / Epidermal |
| Яркость / Brightness | 0,467–0,942 | 0,548–0,880 |
| Контраст / Contrast | 0,066–0,333 | 0,069–0,225 |
| Энтропия / Entropy | 3,626–5,590 | 4,175–5,340 |
| Высокочастотная насыщенность по FFT / High-frequency content (FFT blur) | 23,82–56,48 | 24,07–59,00 |

| Средние значения цветовых каналов / Mean values for color channels | | |
|--|-------------|----------------------------|
| mean_r | 0,622–1,000 | 0,644–1,000 |
| mean_g | 0,455–0,953 | 0,570–0,897 |
| mean_b | 0,411–0,849 | 0,478–0,820 |
| Стандартные отклонения цветовых каналов / Standard deviations for color channels | | |
| std_r | 0,022–0,330 | Не включена / Not included |
| std_g | 0,076–0,330 | Не включена / Not included |
| std_b | 0,090–0,325 | Не включена / Not included |
| Резкость / Sharpness | 0–0,0020 | 0–0,0025 |
| Доля пересвеченных пикселей / Percentage of overexposed pixels | 0–0,300 | 0–0,200 |
| Доля затемненных пикселей / Percentage of underexposed pixels | 0–0,030 | 0–0,010 |

Примечание. FFT (англ. fast Fourier transform) – быстрое преобразование Фурье.

Note. FFT – fast Fourier transform.

Сопоставление двух наборов наглядно демонстрирует, как изменяются нормативные значения параметров при переходе от одного класса меланоцитарных опухолей к другому классу немеланоцитарных опухолей кожи в пределах одной предметной области. Референсный диапазон яркости в меланоцитарном датасете шире (0,467–0,942 против 0,548–0,880) и сдвинут в сторону более низких значений, что отражает большую пигментную вариабельность меланоцитарных образований. Диапазон энтропии, напротив, в эпидермальном датасете уже 4,175–5,340 (против 3,626–5,590) и смещен к более высоким значениям, что соответствует более однородной текстурной архитектуре эпидермальных опухолей.

Эффективность нормализации параметров / Efficiency of parameter normalization

Применение модуля адресной предобработки обеспечило воспроизводимое улучшение всех ключевых метрик диагностической точности во всех трех исследованных конфигурациях ПКЗ (**табл. 7, рис. 6**). Прирост точности составил +0,014–0,017 в абсолютных значениях в зависимости от датасета и программы.

Таблица 7. Метрики диагностической точности до и после применения модуля адресной предобработки фотоизображений

Table 7. Diagnostic accuracy metrics before and after applying the targeted photo image preprocessing module

| | Меланоцитарные фотоизображения / Melanocytic photo images | | Немеланоцитарные фотоизображения / Non- melanocytic photo images |
|--|--|----------------|--|
| Метрика / Metric | Derma Onko Check | Melanoma Check | Derma Onko Check |
| Точность / Accuracy | | | |
| до / before | 0,909 | 0,844 | 0,868 |
| после / after | 0,926 | 0,857 | 0,882 |
| Прирост точности / Accuracy increase | +0,017 | +0,014 | +0,015 |
| Чувствительность / Sensitivity | | | |
| до / before | 0,949 | 0,923 | 0,882 |
| после / after | 0,961 | 0,932 | 0,893 |
| Прирост чувствительности / Sensitivity increase | +0,012 | +0,009 | +0,010 |
| Специфичность / Specificity | | | |
| до / before | 0,901 | 0,827 | 0,860 |
| после / after | 0,919 | 0,842 | 0,877 |
| Прирост специфичности / Specificity increase | +0,019 | +0,015 | +0,017 |

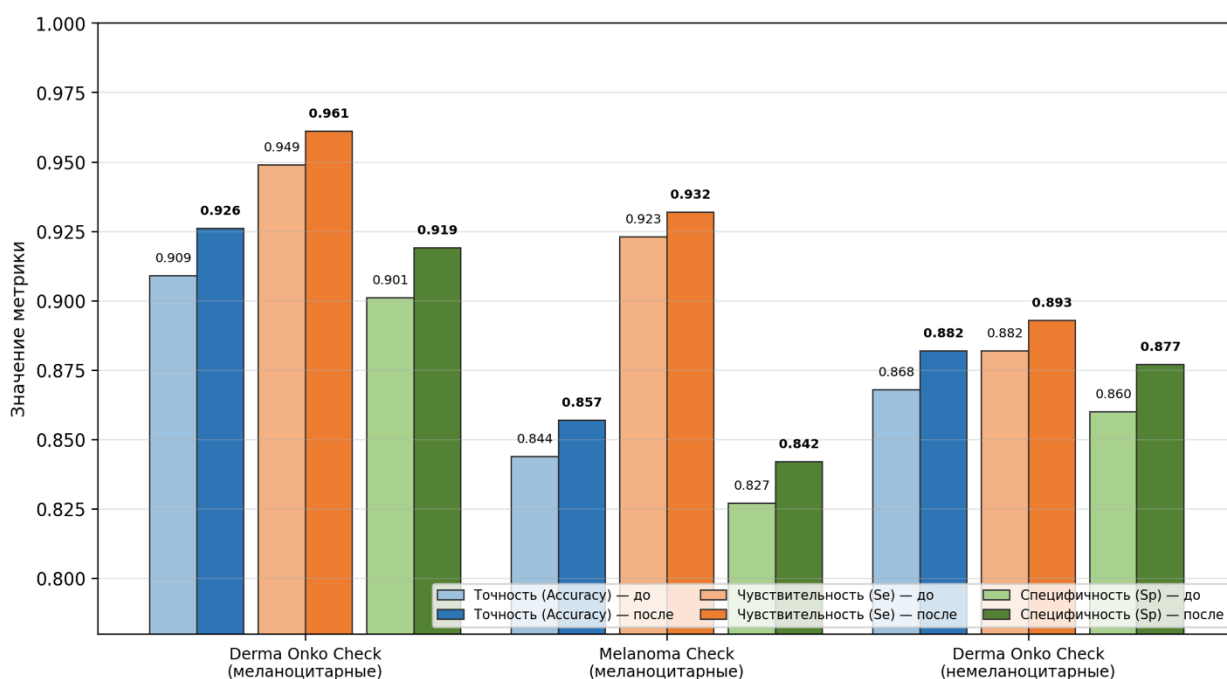


Рисунок 6. Сравнение метрик диагностической точности программ компьютерного зрения до и после применения модуля адресной предобработки фотоизображений

Figure 6. Comparison of diagnostic accuracy metrics of computer vision programs before and after applying the targeted photo image preprocessing module

Наиболее выраженный прирост точности зафиксирован для программы *Derma Onko Check* на меланоцитарном датасете (+0,017), что согласуется с большей распространенностью фотоизображений с отклоняющимися параметрами в данной выборке. Специфичность во всех конфигурациях демонстрировала более существенный прирост (+0,015–0,019) по сравнению с чувствительностью (+0,009–0,012), что свидетельствует о преимущественном снижении доли ЛП заключений. Подобная закономерность согласуется с количественным результатом логистической регрессии, согласно которому доля пересвеченных пикселей входит в число статистически значимых предикторов ошибочного заключения (OR 1,64; $p < 0,05$): устранение доли пересвеченных пикселей и нормализация цветового баланса непосредственно адресуют главную причину ЛП ошибок.

ОБСУЖДЕНИЕ / DISCUSSION

Полученные данные о ведущей роли фотометрических параметров (яркости, контраста, размытости) в формировании ошибочных заключений программ компьютерного зрения соответствуют направлению исследований применения объяснимого искусственного интеллекта в дерматологии, в которых вариабельность условий получения фотоизображений, перепады яркости и контраста, нерезкость, особенности масштаба и поворота кадра, а также артефакты в виде линеек и маркировочных меток, рассматривается как самостоятельный источник снижения диагностической точности сверточных нейронных сетей [9]; в качестве дополнительного фактора систематического смещения обсуждаются особенности цветовой балансировки [10].

Применение методов объяснимого искусственного интеллекта для анализа причин ошибочной классификации становится общепринятым методологическим направлением. Показано, что нейронные сети могут опираться на нерелевантные элементы кадра (в частности, отмечена статистически значимая отрицательная корреляция точности *DeepDerm* с присутствием линеек и маркировочных меток [10]), а систематические смещения, унаследованные от обучающих выборок, описаны как в масштабном анализе 70 работ [11], так и в систематическом обзоре сравнительных исследований [12]. Это согласуется с полученными в настоящей работе методом *Occlusion Sensitivity* данными о смещении внимания модели ИИ на фоновые элементы при ложноотрицательных заключениях и обосновывает обязательную валидацию референсных диапазонов на независимых мультицентровых выборках.

Значимость высокочастотной насыщенности по быстрому преобразованию Фурье как независимого предиктора ошибочной классификации (отношение шансов 3,08; $p < 0,001$) количественно подтверждает описанную в литературе критическую роль качества фокусировки [9] и чувствительность нейронных сетей к малым изменениям параметров входного изображения [12].

Следует отметить, что предложенная в настоящей работе методология, установление количественных референсных диапазонов фотометрических параметров изображения для прогнозирования вероятности ошибочной классификации программами компьютерного зрения, на момент написания статьи не имеет прямых аналогов в доступной литературе.

Предложенная методология реализует переход от констатации факта ошибочной классификации к содержательному анализу ее причин и далее к практическим инструментам коррекции. Принципиальная особенность подхода состоит в том, что источники ошибок раскрываются одновременно с трех сторон: со стороны количественных характеристик входного фотоизображения (статистический анализ метрик качества на одно- и многофакторном уровнях), со стороны логики нейронной сети (методы объяснимого ИИ) и со стороны количественного ранжирования предикторов вероятности ошибки (логистическая регрессия). Подобное многоуровневое представление позволяет дифференцировать ошибки, обусловленные техническими дефектами фотоизображения, от ошибок, обусловленных морфологическим сходством патологических и непатологических образований.

Технические причины ошибочной классификации, выявляемые методологией, включают переэкспозицию (повышенная яркость и доля пересвеченных пикселей), недоэкспозицию (пониженная яркость и доля затемненных пикселей), размытость (сниженная резкость и высокочастотная насыщенность), цветовой дисбаланс (отклонения средних значений и стандартных отклонений каналов R, G, B) и низкую текстурную информативность (сниженная энтропия). Каждая из этих причин может быть устранена соответствующим этапом адресной предобработки, что обеспечивает прямую трассируемость от выявленного отклонения параметра к примененной коррекции и далее к приросту диагностической точности.

Клинические причины ошибочной классификации существенно сложнее технических и связаны с фундаментальной морфологической изменчивостью патологических образований. В дерматологии меланома и диспластический невус могут быть практически неразличимы по макроскопическим признакам, что создает «серые зоны» диагностики даже для опытных врачей-специалистов. Нейронная сеть,

оперирующая исключительно пикселями, лишена клинического контекста (анамнеза, динамики изменений, возможности применения дерматоскопии высокого разрешения), что принципиально ограничивает ее диагностический потенциал в пограничных случаях. Умеренное значение псевдокоэффициента детерминации МакФаддена (0,085) в логистической модели, построенной по фотометрическим и текстурным метрикам, количественно подтверждает данное ограничение: параметры качества фотоизображения объясняют существенную, но не исчерпывающую долю изменчивости в формировании ошибочных заключений, а оставшаяся доля изменчивости отражает действие морфологических факторов, не сводимых к фотометрическим характеристикам.

Применимость методологии в различных областях медицинской визуализации / Applicability of the methodology in various areas of medical imaging

Универсальный характер предложенной методологии определяется тем, что ее этапы не зависят от конкретной предметной области медицинской визуализации. Все девять этапов сохраняют свою последовательность и содержательное наполнение при переходе от дерматологии к офтальмологии, гистологии или ультразвуковой диагностике. Изменяются лишь конкретные значения референсных диапазонов и критических порогов, которые рассчитываются заново для каждой предметной области по характеристикам правильно классифицированных фотоизображений соответствующего домена.

В офтальмологии при анализе фотоизображений глазного дна референсный диапазон яркости закономерно смещен в сторону более низких значений в связи с использованием инфракрасной подсветки и пигментацией глазного дна. Среднее значение красного канала, напротив, стабильно повышено вследствие преобладания красно-оранжевых оттенков сетчатки, поэтому диапазон `mean_r` будет уже сдвинут вверх по сравнению с дерматологическим. Доля пересвеченных пикселей в офтальмологии критична в зоне диска зрительного нерва, где локальная переэкспозиция способна маскировать патологические изменения.

В гистологии при цифровом сканировании микропрепаратов референсные диапазоны параметров определяются стандартизованным протоколом окрашивания и параметрами сканирующей оптики. Принципиальное отличие гистологии от других областей медицинской визуализации состоит в том, что обязательным этапом подготовки обучающих данных служит нормализация по окрашиванию, обеспечивающая согласованность интенсивности и распределения цветовых каналов сканированных препаратов. По этой причине систематические отклонения яркости или контраста, выявляемые методологией в гистологических наборах данных, как правило, не связаны с

условиями получения фотоизображения, а указывают на методологические недостатки обучающей выборки – отсутствие или некорректное применение алгоритмов нормализации окрашивания. В таких случаях практическим следствием анализа становится не нормализация входных данных на этапе инференса, а пересмотр методологии формирования обучающего набора.

В ультразвуковой диагностике, где фотоизображения формируются как черно-белые представления акустического сигнала, цветовые метрики теряют диагностическую значимость, тогда как метрики контраста и текстурной информативности приобретают решающее значение. Высокий уровень спекл-шума (англ. speckle – пятно) требует учета при расчете резкости и высокочастотной насыщенности: соответствующие референсные диапазоны для ультразвукового исследования закономерно отличаются от диапазонов в дерматологии и гистологии.

Практическое значение модулей исключения изображений / Practical significance of image exclusion modules

Закономерным практическим следствием применения методологии является интеграция в медицинские приложения модулей, автоматически исключающих фотоизображения с аномальными значениями параметров. Подобный модуль выполняет три самостоятельные функции. Во-первых, он проверяет соответствие параметров фотоизображения критическим порогам и при их превышении полностью исключает фотоизображение из инференса с уведомлением пользователя. Во-вторых, он проверяет соответствие параметров референсным диапазонам и при выявлении значимых отклонений активирует соответствующие этапы адресной предобработки. В-третьих, он осуществляет логирование выявленных отклонений, что обеспечивает накопление статистики качества фотоизображений в реальных условиях применения.

Принципиальная значимость подобных модулей особенно отчетливо проявляется в телемедицинских сценариях, где возможность запросить повторное фотоизображение с корректными параметрами съемки часто отсутствует. В таких случаях модуль исключения и адресной предобработки выступает не вспомогательным инструментом, а необходимым условием получения достоверного диагностического заключения. Аналогичная ситуация складывается в первичном скрининге, ориентированном на пользователей без специальной подготовки в области медицинской фотографии: неопытный пользователь с большой вероятностью получает фотоизображение с одним или несколькими фотометрическими дефектами, и в отсутствие модуля адресной предобработки такие фотоизображения генерировали бы ненадежные диагностические заключения.

Ограничения / Limitations

Ограничения настоящей методологии определяются несколькими принципиальными обстоятельствами.

Во-первых, корректность установленных референсных диапазонов зависит от объема и репрезентативности выборки правильно классифицированных фотоизображений: при малом объеме выборки или ее систематическом смещении референсные диапазоны могут оказаться непредставительными.

Во-вторых, методология ориентирована на анализ ошибок, обусловленных параметрами входных данных, и не охватывает в полной мере ошибок, связанных с ограничениями архитектуры модели или с недостаточным разнообразием обучающей выборки в части морфологических вариантов патологии. На это указывает умеренное значение псевдокоэффициента детерминации логистической модели (0,085).

В-третьих, при переносе модели ИИ между регистрирующими устройствами (например, между разными моделями смартфонов или фундус-камер или ретинальной камеры для фотографирования глазного дна) референсные диапазоны должны переустанавливаться, что предполагает наличие представительной выборки фотоизображений для каждого устройства.

Также принципиальным методологическим ограничением предложенного подхода является потенциальный риск циклического вывода. Предложенные референсные диапазоны параметров фотоизображений рассчитываются на основе уже классифицированных моделью ИИ изображений, отнесенных к категориям истинно положительных и истинно отрицательных результатов. Если обучающая выборка самой модели ИИ была систематически смещена в сторону определенных условий съемки (например, по освещенности, фоновой текстуре или цветовому балансу), то «правильно классифицированные» изображения могут отражать не только морфологические признаки патологии, но и артефактные корреляции с условиями фотографирования. В таком случае установленные референсные диапазоны закрепляют не диагностически значимые характеристики, а особенности обучающей выборки, что воспроизводит исходное смещение модели.

Для частичного контроля данного риска в методологии предусмотрен этап локализации областей интереса методами объяснимого ИИ (Score-CAM и Occlusion Sensitivity). Преобладание активации тепловых карт в зоне самого новообразования, а не на фоновой коже или артефактах кадра, при истинно положительных заключениях свидетельствует в пользу морфологической природы классификационных решений модели

и снижает вероятность артефактной корреляции. Однако для окончательного исключения возможности циклического вывода необходима валидация предложенных референсных диапазонов на независимых выборках, собранных в стандартизированных условиях съемки, либо на фотоизображениях, прошедших полную фотометрическую нормализацию по яркости, контрасту и цветовому балансу. Указанное обстоятельство относится к ограничениям настоящей методологии и должно учитываться при ее практическом применении.

Указанные ограничения не отменяют основного результата работы. Универсальная структура методологии и предложенный набор фотометрических и текстурных метрик сохраняют свою применимость при расширении выборки и переходе к новым предметным областям медицинской визуализации, что позволяет рассматривать предложенный подход как методологическую основу для дальнейших исследований ошибочной классификации фотоизображений ПКЗ на основе моделей ИИ.

ЗАКЛЮЧЕНИЕ / CONCLUSION

Предложенная обобщенная методология анализа причин ошибочной классификации фотоизображений ПКЗ на основе моделей ИИ позволяет в систематической форме раскрывать источники ошибочных заключений, устанавливать референсные диапазоны фотометрических и текстурных параметров для конкретной предметной области медицинской визуализации и формировать критерии исключения фотоизображений с аномальными значениями метрик. Девять последовательных этапов методологии обеспечивают переход от исходной выборки фотоизображений к практическому результату в виде интегрируемого в медицинское приложение модуля исключения и адресной предобработки.

Принципиальная особенность обновленной методологии состоит в интеграции этапа многомерного статистического анализа, дополняющего классический однофакторный анализ. Двухфакторный ANOVA обеспечивает одновременный учет категории заключения и применяемой программы ПКЗ, MANOVA верифицирует установленные закономерности на уровне совокупного вектора метрик, а множественная логистическая регрессия с контролем мультиколлинеарности позволяет количественно ранжировать вклад отдельных параметров в вероятность ошибочного заключения. Принципиально важно, что MANOVA не заменяет, а дополняет и подтверждает результаты однофакторного ANOVA: совпадение значимых факторов в обеих моделях служит независимым подтверждением воспроизводимости установленных закономерностей.

Выполненная апробация методологии на наборе из 460 фотоизображений меланоцитарных опухолей кожи и независимом наборе немеланоцитарных опухолей кожи показала ее работоспособность на практике. Установлены статистически значимые межгрупповые различия параметров фотоизображений ($F=13,50-39,31$ и $p<0,001$ по основным метрикам однофакторного ANOVA; $F=5,41-72,29$ и $p<0,001$ для категории заключения по двухфакторному ANOVA), верифицирована независимость закономерностей от программы ПКЗ (лямбда Уилкса 0,998; $F=0,138$; $p=0,995$ по MANOVA), установлены ключевые независимые предикторы ошибочного заключения (высокочастотная насыщенность по FFT: $OR=3,08$, $p<0,001$; резкость: $OR=0,31$, $p<0,01$; доля пересвеченных пикселей: $OR=1,64$, $p<0,05$). Сформулированы референсные диапазоны яркости (0,467–0,942 для меланоцитарного датасета и 0,548–0,880 для эпидермального), контраста (0,066–0,333 и 0,069–0,225 соответственно), энтропии (3,626–5,590 и 4,175–5,340 соответственно) и цветовых характеристик. Установлены критические пороги (доля пересвеченных или затемненных пикселей более 55%, яркость менее 0,07 или более 0,97, контраст менее 0,008), при превышении которых фотоизображение исключается из инференса. Применение модуля адресной предобработки обеспечило прирост диагностической точности на +0,014–0,017 в абсолютных значениях во всех исследованных конфигурациях ПКЗ при преимущественном повышении специфичности (+0,015–0,019).

Методология носит универсальный характер. Ее применение в офтальмологии, гистологии, ультразвуковой диагностике и других областях медицинской визуализации требует лишь переустановки референсных диапазонов и критических порогов с учетом специфических условий получения фотоизображений в каждой предметной области. Принципиальный методологический тезис состоит в том, что нормативные значения параметров фотоизображения должны устанавливаться отдельно для каждой области медицинской визуализации и для каждой модели ПКЗ, поскольку перенос диапазонов между областями методологически некорректен и закономерно приводит к росту доли ошибочных заключений.

Реализация в составе медицинских приложений модулей исключения фотоизображений с аномальными значениями метрик и адресной предобработки фотоизображений с отклоняющимися параметрами позволяет существенно повысить клиническую надежность ПКЗ на основе ИИ в реальных условиях применения, особенно при работе с фотоизображениями субоптимального качества и в сценариях телемедицинской диагностики, где возможность повторного фотодокументирования ограничена.

| ИНФОРМАЦИЯ О СТАТЬЕ | ARTICLE INFORMATION |
|---|---|
| <p>Поступила: 29.04.2026 В доработанном виде: 01.06.2026 Принята к печати: 05.06.2026 Опубликована онлайн: 11.06.2026</p> | <p>Received: 29.04.2026 Revision received: 01.06.2026 Accepted: 05.06.2026 Published online: 11.06.2026</p> |
| Вклад авторов | Authors' contribution |
| <p>Ламоткин А.И.: разработка методов и процедур; проведение экспериментов, сбор данных; управление данными: подготовка, аннотирование, хранение; проверка результатов, воспроизводимости; статистический или теоретический анализ данных; создание программного обеспечения, скриптов, алгоритмов; написание первоначального текста статьи; создание графиков, схем, иллюстраций. Корабельников Д.И.: формулирование идеи, гипотезы, цели исследования; разработка методов и процедур; проведение экспериментов, сбор данных; проверка результатов, воспроизводимости; научное руководство проектом; организационное управление проектом; редактирование, доработка текста. Авторы прочитали и утвердили окончательный вариант рукописи</p> | <p>Lamotkin A.I.: development of methods and procedures; conducting experiments, data collection; data management: preparation, annotation, storage; verification of results and reproducibility; statistical or theoretical analysis of data; creation of software, scripts, algorithms; drafting of the initial text of the article; creation of graphs, diagrams and illustrations. Korabelnikov D.I.: formulation of the idea, hypotheses and research objectives; development of methods and procedures; conducting experiments and data collection; verification of results and reproducibility; scientific supervision of the project; organisational management of the project; editing and finalising the text. The authors have read and approved the final version of the manuscript</p> |
| Конфликт интересов | Conflict of interests |
| Авторы заявляют об отсутствии конфликта интересов | The authors declare no conflict of interests |
| Финансирование | Funding |
| Авторы заявляют об отсутствии финансовой поддержки | The authors declare no funding |
| Этические аспекты | Ethics declarations |
| Неприменимо | Not applicable |
| Раскрытие данных | Data sharing |
| Первичные данные могут быть предоставлены по обоснованному запросу автору, отвечающему за корреспонденцию | Raw data could be provided upon reasonable request to the corresponding author |
| Комментарий издателя | Publisher's note |
| Содержащиеся в этой публикации утверждения, мнения и данные были созданы ее авторами, а не издательством ИРБИС (ООО «ИРБИС»). Издательство снимает с себя ответственность за любой ущерб, нанесенный людям или имуществу в результате использования любых идей, методов, инструкций или препаратов, упомянутых в публикации | The statements, opinions, and data contained in this publication were generated by the authors and not by IRBIS Publishing (IRBIS LLC). IRBIS LLC disclaims any responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred in the content |
| Права и полномочия | Rights and permissions |
| © 2026 А.И. Ламоткин, Д.И. Корабельников; ООО «ИРБИС» Статья в открытом доступе по лицензии CC BY-NC-SA (https://creativecommons.org/licenses/by-nc-sa/4.0/) | © 2026 A.I. Lamotkin, D.I. Korabelnikov. Publishing services by IRBIS LLC This is an open access article under CC BY-NC-SA license (https://creativecommons.org/licenses/by-nc-sa/4.0/) |

ЛИТЕРАТУРА / REFERENCES

1. Ламоткин А.И., Корабельников Д.И., Олисова О.Ю., Ламоткин И.А. Эффективность предварительной дифференциальной диагностики доброкачественных и злокачественных новообразований кожи с помощью программы искусственного интеллекта Derma Onko Check. *ФАРМАКОЭКОНОМИКА. Современная фармакоэкономика и фармакоэпидемиология.* 2025; 18 (2): 261–70. <https://doi.org/10.17749/2070-4909/farmakoeconomika.2025.294>.

Lamotkin A.I., Korabelnikov D.I., Olisova O.Yu., Lamotkin I.A. Effectiveness of preliminary differential diagnosis of benign and malignant skin neoplasms using the Derma Onko Check artificial intelligence program. *FARMAKOEKONOMIKA. Sovremennaya farmakoeconomika i farmakoepidemiologiya / FARMAKOEKONOMIKA. Modern Pharmacoeconomics and Pharmacoepidemiology*. 2025; 18 (2): 261–70 (in Russ.). <https://doi.org/10.17749/2070-4909/farmakoeconomika.2025.294>.

2. Nadour N., Duguet T., Zahedi S., et al. Diagnostic accuracy of artificial intelligence compared to family physicians and dermatologists for skin conditions: a systematic review and meta-analysis. *BMC Prim Care*. 2025; 26 (1): 384. <https://doi.org/10.1186/s12875-025-03073-9>.

3. Zapata M.A., Royo-Fibla D., Font O., et al. Artificial intelligence to identify retinal fundus images, quality validation, laterality evaluation, macular degeneration, and suspected glaucoma. *Clin Ophthalmol*. 2020; 14: 419–29. <https://doi.org/10.2147/OPTH.S235751>.

4. Prasad P., Khair A.M.B., Saeed M., Shetty N. Artificial intelligence in histopathology. *J Pharm Bioallied Sci*. 2024; 16 (Suppl. 5): S4226–9. https://doi.org/10.4103/jpbs.jpbs_727_24.

5. Khalaf K., Terrin M., Jovani M., et al. A comprehensive guide to artificial intelligence in endoscopic ultrasound. *J Clin Med*. 2023; 12 (11): 3757. <https://doi.org/10.3390/jcm12113757>.

6. Qamar T., Bawany N.Z. Understanding the black-box: towards interpretable and reliable deep learning models. *Peer J Comput Sci*. 2023; 9: e1629. <https://doi.org/10.7717/peerj-cs.1629>.

7. Evans H., Snead D. Understanding the errors made by artificial intelligence algorithms in histopathology in terms of patient impact. *NPJ Digit Med*. 2024; 7 (1): 89. <https://doi.org/10.1038/s41746-024-01093-w>.

8. Корабельников Д.И., Ламоткин А.И. Анализ причин ошибочной классификации изображений немеланоцитарных опухолей кожи программами на основе технологий искусственного интеллекта. *Сибирский журнал клинической и экспериментальной медицины*. 2026; 41 (1): 221–31. <https://doi.org/10.29001/2073-8552-2026-2881>.

Korabelnikov D.I., Lamotkin A.I. Analysis of the reasons of misclassification of non-melanocytic skin tumors by artificial intelligence based programs. *Siberian Journal of Clinical and Experimental Medicine*. 2026; 41 (1): 221–31 (in Russ.). <https://doi.org/10.29001/2073-8552-2026-2881>.

9. Daneshjou R., Vodrahalli K., Novoa R.A., et al. Disparities in dermatology AI performance on a diverse, curated clinical image set. *Sci Adv.* 2022; 8 (32): eabq6147. <https://doi.org/10.1126/sciadv.abq6147>.

11. Daneshjou R., Smith M.P., Sun M.D., et al. Lack of transparency and potential bias in artificial intelligence data sets and algorithms: a scoping review. *JAMA Dermatol.* 2021; 157 (11): 1362–9. <https://doi.org/10.1001/jamadermatol.2021.3129>.

10. Liopyris K., Gregoriou S., Dias J., Stratigos A.J. Artificial Intelligence in dermatology: challenges and perspectives. *Dermatol Ther.* 2022; 12 (12): 2637–51. <https://doi.org/10.1007/s13555-022-00833-8>.

12. Haggemüller S., Maron R.C., Hekler A., et al. Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts. *Eur J Cancer.* 2021; 156: 202–16. <https://doi.org/10.1016/j.ejca.2021.06.049>.

Сведения об авторах / About the authors

Ламоткин Андрей Игоревич, к.м.н. / *Andrey I. Lamotkin*, PhD – ORCID: <https://orcid.org/0000-0001-7930-6018>. eLibrary SPIN-code: 4170-7782. E-mail: lamotkin.an@mail.ru.

Корабельников Даниил Иванович, к.м.н., доцент / *Daniil I. Korabelnikov*, PhD, Assoc. Prof. – ORCID: <https://orcid.org/0000-0002-0459-0488>. eLibrary SPIN-code: 7380-7790.